SURVEY ON HANDWRITTEN DEVNAGARI CHARACTER RECOGNITION

Aditi Nene [1] | Aishawarya Palkar [1] | Megha Nagarhalli [1] | Richa Misri [1] | Prof. Shivraj Kone [1]

[1] Department of Computer Engineering, RMDSinhgad School of Engineering.

ABSTRACT

Character Recognition by machines is an innovative way by which the dependence on manpower is reduced. Character recognition provides a reliable alternative of converting manual text into digitized format. Now-a-days, as technology becomes integral part of human life, many applications have enabled the incorporation of English OCR for real time inputs. The advantages that the English alphabet has is its simplicity offered by less number of lettersi.e. 26 and easier classification due to the concept of lowercase and upper case. If we consider Devnagari script in this scenario, we will come across myriad hurdles because this script lacks the simplicity of English. The concept of fused letters, modifiers, shirorekha and spitting similarities in some letters make recognition difficult. Also, character recognition for handwritten text is far more complex than that for machine printed characters. This is because of the versatility and different writing techniques adopted by people. The direction of strokes, pressure applied on writing equipments, quality of writing equipment and the mentality of the writer itself highly affects the written text. These problems when combined with the intricate details of Devnagari script, the complications in constructing a HCR of this script are increased. The proposed system focuses on these two issues by adopting Hough transform for detecting features from lines and curves. Further, for classification, SVM is used. These two methods when combined provide high accuracy which is up to 90%. Prior to these techniques, pre-processing of characters is done to ensure accurate classification. This system is highly useful as it can be used for automation of various services like postal, rail etc.

I. INTRODUCTION

With the technological development aiming to bridge the gap between humans and machines, research is concentrated in areas like text and pattern recognition. While building any OCR ,there are three basic steps which the input goes through:-

i. Preprocessing: In this step, various functions like curve smoothening, thinning, darkening and binarisation are performed on the input character image so as to make it more readable.

ii. Feature extraction: In this step, the character image obtained from preprocessing step is analysed for collecting unique features from it, like presence of loops or knots. These features are useful for mapping the input to output.

iii. Classification: This is the last step in which features extracted from the second step are maped to various classes of letters. From the similarities evident from collected features, recognition is done.

As handwriting is the most primitive form of information storage and communication, enabling applications and websites to accept handwriting as input will be of great usefulness. Automation for preserving historic texts and deciphering them, automation in fields like postal systems, form filling systems and the simple advantage of storage in a digital format are all possible due to handwritten character recognition. Considering the scenarios in India, where still many documents are still stored in manual format in regional languages like Hindi, Marathi etc. an OCR for Devnagari script is needed.

The Devnagari script has 36 consonants and 13 vowels. The consonants are given modifiers to incorporate the effects of vowels into them. Also, there is a concept of fused letters in this script in which two consonants can be merged. These fused letters are called jodakshara. Thus, it is obvious that it is indeed very challenging to construct an OCR for Devnagari script.
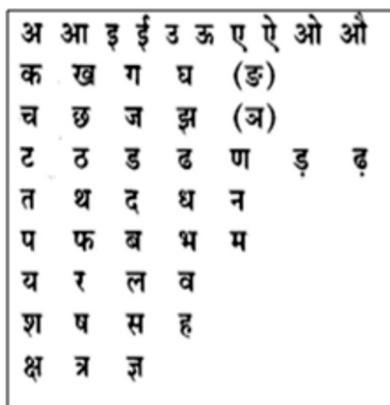

अ आ इ ई उ ऊ ए ऐ ओ औ
क ख ग घ (ङ)
च छ ज झ (अ)
ट ठ ड ढ ण ड़ ढ़
त थ द ध न
प फ ब भ म
य र ल व
श ष स ह
क्ष त्र ज्ञ

Fig 1.Devnagari alphabet

II. LITERATURE REVIEW

It includes offline character recognition method which recognizes handwritten scripts and shows it in typed character format by using ASCII. This system takes handwritten text image as an input and then it isolates lines, words, characters sequentially by using region labeling. For text recognition, templates matching technique was used which finds location of sub-image in input image. After location retrieval, the text is in type character format. Phase correlation provides straight forward estimation of rigid translational motion between two images which is based on the well known Fourier Shift property. The phase correlation is used because it gives high accuracy.

Template matching is used in image processing and computer vision but it has problem if the image has presence of extreme noise [1].For devanagari character recognition, it provides a systematic study of segmentation methods. This system is not only directly concern to the character but also words, phrases and even the complete documents. For the character recognition, Hidden Markova Model, neural networks and their combination are used as powerful tools. In character recognition, for reliability segmentation and classification have to be treated in an integrated manner to obtain more accuracy in complex cases [2].The algorithm which is used in this research for character recognition, it first segments the image containing devanagari text fed to the software into lines, words and characters. The characters obtained are then brought to a standard size. Here Kohonen Neural network based recognizer recognizes the text by character and gives the output in Unicode format. To support quick recognition the network has been designed with no hidden layer. Apart, from text recognition from an image, it can also recognize character drawn using a mouse which leads to keyboard computer interaction. The technique uses java having accuracy as 90.26% and 83.33 % for machine printed and mouse driven characters [3].This paper makes use of Canny edge detection technique and artificial neural network for handwritten hindi character recognition. The steps involved for handwritten hindi character recognition system are :

1. Scanning.
2. Preprocessing.
3. Segmentation.
4. Canny operator.
5. Distance transformation.
6. Feature extraction.
7. Feed back and propogation of artificial neural network for recognition.

The scope of the proposed system is limited to simple character recognition [4].Rectangle histogram oriented gradient representation for the extraction of feature is used. Few simple arithmetic operations per image pixel are used in algorithm for real time application. This paper uses SVM (Support Vector Machine) and FFANN (Feed Forward Artificial Neural Network) classification technique as it is more efficienti.e. effective with increased processing speed and accuracy. A multilayer FFANN is used with 10 hidden layer for the classification purpose. SVM is capable of learning to achieve good generalization error free recognition on their handwritten character dataset. The concept of SVM resolve a classify the hyper plane of SVM in feature space which are non-linearly connected to the input space [5].Particle swarm optimization and support vector machine technique are implemented. An android phone is used for taking input of character as

image and MATLAB software is use for showing the recognized devanagari character. PHP language is used as mediator for connecting MATLAB and android devices. Image processing techniques are performed on the featured character in MATLAB for character classification. PSO algorithm is applied for character classification from input feature. The particle swarm optimization technique has accuracy of 90% [6].The SVM and ANN classification method are applied on handwritten devanagari character. After preprocessing the character image, shadow feature, chain code histogram feature, view based feature and congest run features are extracted. These features are then fed to Neural classifier and support vector machine for classification of characters. In neural classifier, three ways of combining decision of 4 MLP's are designed for four different features. SVM in its elementary form is used for binary classification. It may however be extended to multiclass problem using the one-against- one approach[7].The challenges involved in Indian postal system automation with is presented in paper along with case study. This paper shows the existing research literature support available for postal automation system with a case study of pune city [8].The system deals with building of grid based method which is combination of grid based approach used in feature extraction, for individual with respect to image centroidis computed. In combination of image centroid as well as zone centroid which gives feature vector of size 2*n features. This feature vector is presented to feed forward neural network for recognition. The method used in this paper does the segmentation of handwritten character and recognition using neural network. HCR works in steps as preprocessing, segmentation, feature extraction and recognition using neural network [9]. The paper conveys a use of regular expression and demonstrates the effective use of regular expressions that can bind to facilitate more efficient & more effective character recognition. Regular expressions are represented as a character strings. In proposed method, the characters are first identified with regular expression matching and if they do not match with any pattern they are passed to minimum edit distance filter [10].Handwritten devanagari scripts recognition system using neural network is represented. A diagonal feature extraction schema for the recognition of offline handwritten character is used. For performing classification and recognition task an artificial neural network is used as backend. After extracting feature, character recognize image in which extracted feature are converted into chromosome string. The fitness function is used in the recognition step to find difference how unknown character and chromosome which is stored in database [11].Optical character recognition involves preprocessing, segmentation, feature extraction, classification and script recognition. Skewed algorithm is used for the images which get tilted during scanning. For recognition the individual character is taken as input, here the document is segmented into lines, words and characters. Gabor filter is used for feature extraction of characters[12].The template matching algorithm is used for devanagari script. Characters are taken in OCR from document image. The scope of this proposed system is limited to the recognition of a single character. To convert the human readable documents into computer process able form, OCR systems are developed. In this, the scanned image of machine printed or handwritten text, numerals, letters and symbols into a computer process able format such as ASCII [13].An image is taken as input and then that is used to convert it as an output to OCR. This paper uses the idea of creating system for recognizing character using neural network. Here the user writes the character which is then processed using image processing and then processed image is given and trained using neural network. Sobeltechnique is used to reduce the noise and give the proper normalized image. The normalized image is then given for feature extraction where characters are uniquely identified using neural networks and then final output will be displayed after postprocessing [14].The main focus of paper is on offline recognition of handwritten devanagari character using segmentation and artificial neural network. The whole process includes segmentation of characters into lines, words and characters and then recognization through feed forward neural network. The system is capable of recognizing handwritten characters or symbols with the help of neural network [15].

III. PROBLEMS IDENTIFIED

Optical character recognition for Devnagari script is highly affected by the database that the system is using. To obtain all the combinations letters with numerals along with modifiers and half forms is a strenuous job. Also, when we consider handwritten text, the handwritings need to be collected from a sufficient number of individuals.

Another problem is that the precision that the user requires. If the system is very sensitive, classification will become complicated. Further, the device reliability matters.

IV. CONCLUSION AND FUTURE WORK

In the world that we live in, time is the most important factor affecting human lives. In order to, cut off time required to convert manual writing into digital ones we need software that would do this work to replace a human being. As devnagari script is adopted by many languages in India, an application that digitizes devnagari written language is needed. Going into further detail, many organizations and institutions are shifting towards paperless working and need to convert user's handwritten manual documents into digital documents. Ours proposed system is ideal for situation like these. Considering all the complications that comes into picture with regard to handwritten devnagari OCR, the proposed system strives to achieve the maximum accuracy.

As most of the noise and pixel irregularities are removed in the pre-processing stage, accuracy is increased. In the feature extraction stage which is the most important step of an OCR, detection of feature from lines and curves outputs to more précised collection of unique features. Considering the techniques that have been implemented over the years, it is observed that the combinations of techniques used affect the outcome. The OCR depending upon the end user and the resources available appropriate techniques should be used at each stage to increase accuracy.

As per survey, we find the use of HTDCC most suited at the feature extraction stage and SVM most suited for classification. The implementation of these two techniques is best suited for our project.

VI. REFERENCES

[1] Poonam M. Ingle, P.P. Gumaste. Handwritten Devnagari Script Recognition by using Phase Correlation; 2014

[2] Mr. Kuldeep P. Pawar, Mr. Digvijay J. Pawar, Mr. Yashwant S. Jagadale. A systematic approach to Devnagari Character Recognition method;2015

[3] Pulkit Goyal, Sapan Diwakar, Anupam Agrawal. Devnagari Character Recognition towards Natural Human-Computer Interaction;2010

[4] Tanuja K, Usha Kumari V and Sushma T. M. Handwritten Hindi Character Recognition System using Edge Detection and Neural Network;2015

[5] Parshuram M. Kamble, Ravinda S. Hegadib. Handwritten Marathi Character Recognition using R-HOG Feature;2015

[6] Prashant M. Kakde, Dr. S. M. Gulhane. A comparative analysis of particle swarm optimization and support vector machines for devnagari character recognition: An Android Application.;2016

[7] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, L. Malik, M. Kundu and D. K. Basu. Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition;2010

[8] Ankita S. Wanchoo, Preet iYadav, Alwin Anuse. A survey on Devnagari Character Recognition for Indian Postal System Automation;2016

[9] Ms. Seema A. Dongare, Prof. Dhananjay B. Kshirsagar, Ms. Snehal V. Waghchaure. Handwritten Devnagari Character Recognition using Neural Network;2014

[10] Dr. P. S, Deshpande, Latesh Malik, Sandhya Arora. Fine Classification and Recognition of Handwritten Devnagari Characters with Regular Expressions and Minimum Edit Distance Method;2008

[11] VedPrakash Agnihotri. Offline Handwritten Devnagari Script Recognition

[12] Asmita Kunkari. Optical Character Recognition System For Devnagari Script;2016

[13] Shruti Agarwal, Dr. Naveen Hemarjani. Offline Handwritten Character Recognition with Devnagari Script;2013

[14] Ankita Karia, Sonali Sharma, Reevon Rodrigues, Maitreya Save. Character Recognition (Devnagari Script)

[15] Mitrakshi B. Patil, Vaibhav Narawade. Recognition of Handwritten Devnagari Characters through Segmentation and Artificial neural networks;2012